**Boat Sales Data**

**Data Source:** The data for this dataset is sourced from Kaggle.com revolves around a boat sales website where individuals sell both new and used boats. This website serves as a platform for boat owners to list their boats for sale, providing a diverse and dynamic source of information related to boat sales. The data source is directly tied to the boat sales market, making it highly relevant for marketing and trend analysis in the boating industry.

**Data Collection:** The data collection process for this dataset is continuous and ongoing, with updates expected on an annual basis. This regularity in updates ensures that the dataset remains current and reflective of changing market conditions. The marketing team uses this data to create a weekly newsletter for boat owners, helping them enhance the visibility of their listings and stay informed about market trends. The data collection process likely involves web scraping or API access to gather information from the boat sales website.

**Contents:** This dataset comprises 10 columns of information, focusing primarily on boat popularity. Key attributes include details about the boat's year of manufacture, its type, price, and various other relevant features. The dataset appears to be rich in boat-specific information, providing a comprehensive overview of boats available for sale on the website. Such details are valuable for understanding market trends and assisting boat sellers in making informed decisions.

**Limitations and Ethics:** While this dataset offers valuable insights into the boat sales market, it may have certain limitations. One potential limitation is the accuracy and completeness of the data, as it relies on user-generated content on the website. There could be errors or discrepancies in the listings, affecting the dataset's reliability. Additionally, ethical considerations such as data privacy and consent should be addressed, ensuring that the personal information of boat owners and sellers is handled responsibly and in compliance with relevant regulations.

**Data Relevancy: This data was provided in the CareerFoundry Project Brief. This data collection fits the project's requirements since it is open source, contains geographical objects such as country and city, and comprises 9888 rows and 10 columns. This dataset is highly relevant for my analysis, especially given my interest in the subject of sales. The data was picked because it is about sales and marketing and trends, and I am interested in sales and marketing strategy, sales analysis, and product success planning. By selecting this dataset to work on and learn from, I will undoubtedly develop more confidence and improve my abilities in this subject.**

**Data Profile**

There are 9888 rows and 10 columns in original dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Description** | **Time -Variant/-Invariant** | **Structured/**  **Unstructured** | **Qualitative/**  **Quantitative** | **Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous** |
| **Price** | Price of the boat in different currencies. | Time -Variant | Structured | Quantitative | Countinuous |
| **Boat Type** | The type of boat e.g. fishing, sport, yatch etc. | Time -Invariant | Structured | Qualitative | Nominal |
| **Manufacturer** | Name of the manufacture who made the boat | Time-Invariant | Structured | Qualitative | Nominal |
| **Type** | New boat from stock / Used boat | Time-Invariant | Structured | Qualitative | Nominal |
| **Year Built** | Year of the boat was built | Time-Variant | Structured | Quantitative | Countinuous |
| **Length** | Length size of the boat | Time-Invariant | Structured | Quantitative | Countinuous |
| **Width** | Width size of the boat | Time-Invariant | Structured | Quantitative | Nominal |
| **Material** | Material of the boat was made | Time-Invariant | Structured | Qualitative | Nominal |
| **Location** | Country and city of the boat | Time-Invariant | Structured | Qualitative | Nominal |
| **Number of views last 7 days** | Number of people view the boat on the website | Time-Variant | Structured | Quantitative | Discrete |

**Data Integrity**

**Data Accuracy (numeric columns)**

A table with numbers and lines

Description automatically generated

**Cleaning Data**

**Column Derivations and aggregations**

|  |  |  |
| --- | --- | --- |
| New column | Column/s it was derived from | Conditions |
| EUR\_price | Price | Convert price rate in Euros.  'CHF \*1.04',  'DKK \*0.13',  'Åƒ \*1.16' |
| Country | Location | Separate Country name |
| City | Location | Separate City name |

**Wrangling steps**

|  |  |  |  |
| --- | --- | --- | --- |
| Columns dropped | Columns renamed | Columns' type changed | Comment/Reason |
| Location |  |  | Split the location column into 2 columns Country and City. |
| Price |  |  | Convert all prices into Euros |

**Consistency check**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Missing values** | **Missing values treatment** | **Duplicates** |
| **Width** | 56 | Replace the missing values with Median as the data skewed to right | No duplicates found |
| **Length** | 9 | Replace the missing values with Median as the data skewed to right | No duplicates found |
| **Price** |  |  | No duplicates found |
| **Boat Type** |  |  | No duplicates found |
| **Manufacturer** | 1338 | Missing values converted to ‘none’ | No duplicates found |
| **Type** | 6 | Filled missing values with mode value | No duplicates found |
| **Year Built** | 551 | This column had ‘0’ values which was droped. | No duplicates found |
| **Material** | 17749 | Missing values converted to ‘none’ | No duplicates found |
| **Country** | 36 | Dropped the missing values | No duplicates found |
| **Number of views last 7 days** |  |  | No duplicates found |
| **City** | 36 | Dropped the missing values |  |

**Questions to be Explore:**

1. Which regions in Europe are boats looked at?
2. characteristics of the most-viewed boat listings in the last 7 days.
3. Is it the most expensive boat that gets the most views?
4. Are there common features among the most-viewed boats?